Analysis and Recommendations for the DTIC Non-Print Collection

Connie J. Wiley
Information Science and Technology Directorate
DTIC-E
Defense Technical Information Center
November 25, 2002

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Nov 2002	Final	-

Title and Subtitle Analysis and Recommendations for the DTIC Non-Print Collection	**Contract Number**
	Grant Number
	Program Element Number
Author(s) Connie J. Wiley	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Defense Technical Information Center DTIC-E 8725 John J. Kingman Road Ft. Belvoir, VA 22060	**Performing Organization Report Number** DTIC-E-TR-2003/02
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes

Abstract
This study examined the DTIC non-print collection to determine its composition and potential preservation issues. It determined how DTIC policies and practices affect long-term access, and recommended a course of action for future collection access. The non-print collection has very different requirements than DTICs traditional technical reports collection. Text objects (technical reports) do not require the individualized attention and resources of the non-print collection. This study presents short and long-term recommendations for balancing customer need for long-term access with DTIC's need to make the most of finite resources. Rather than continuing to support all contributions equally regardless of long-term value, it suggests that DTIC develop a plan to emphasize those digital objects already in its collection and that merit long-term accessibility; and concentrate on development of a future collection that is preservation-ready. By guiding the development of a future collection, DTIC will continue to provide useful content and services to customers.

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** SAR

Number of Pages
46

Executive Summary

This study was intended to accomplish three goals: 1) examine the DTIC non-print collection to determine its composition and potential preservation issues; 2) determine how DTIC policies and practices affect long-term access; and 3) recommend a course of action for future collection access. This study was not intended to address the text/print collection or the actual science of preservation. In terms of physical format, maintenance needs and long-term accessibility, the non-print collection has very different requirements than DTIC's traditional technical reports collection. Although text objects (technical reports) comprise the largest part of the DTIC collection, they are in or can be transferred to one uniform format. As a collection, they do not require the individualized attention and resources of the media collection.

It is reasonable to equate effective, long-term accessibility with preservation. Preservation, like any technical discipline, is an evolving science. It can never be finished; whatever is accomplished today will change as new technology is introduced. Preservation requires a continuous organizational commitment. Once preservation action is taken, repeated attention must be paid to an object over its lifespan. This study suggests balancing the needs of customers for long-term access with the well-being of DTIC by outlining ways to make the most of finite resources. Rather than continuing past practice of supporting all contributions equally and as a separate collection, regardless of long-term value, it suggests that DTIC revise the way it thinks about this collection. Emphasis should shift to those objects that truly merit long-term accessibility.

Short and long-term recommendations address specific issues. Short-term suggestions are to 1) evaluate the collection, item by item, to judge what should be retained; 2) determine how each object should be retained; 3) enhance current metadata to permit better access; 4) provide content to customers in multiple formats as well as online; 5) simplify the collection by de-selecting content in duplicative file formats; and 6) insure that collection managers have access to tools and training specific to the discipline. Long-term recommendations are to 1) better understand customers' needs; 2) strengthen contributor guidelines; 3) more clearly define what is collected; 4) develop an overall preservation policy; and 5) watch the technology.

This paper suggests that DTIC can derive greater return on investment by concentrating on future contributions rather than taking extensive preservation action on the entire retrospective collection. By creating a collection that is lean, mean and preservation-ready, and by guiding the

development of a future collection, DTIC will continue to provide useful information and services to customers.

SECTION I: APPROACH

Part I: ANALYSIS

A. Study Goal

This study was intended to accomplish three goals:

- Examine the non-print collection to determine its composition and potential preservation concerns;
- Determine how DTIC policies and practices affect access to that collection over the long-term; and
- Recommend a course of action for long-term access to that collection.

One of the issues this paper attempts to address is the wisdom of maintaining a separate non-print collection. The collection was begun in 1991, when contributors and customers became increasingly able to produce and use content such as software and data files on non-paper media. As a natural extension of its mission, DTIC took on the job of reproducing and distributing products, in part to support these contributors who did not have the resources or facilities to do so.

Since that time, the growth of the Web and other technologies has surpassed removable media as a method of sharing information. This paper attempts to suggest ways to rethink how this collection is handled and how a wider group of customers could be supported by it within the constraints of present resources.

If preservation of the content is attempted, it will require commitment to constant change over the life-cycle of an object. Guaranteeing permanent access to all objects, regardless of content, age, customer interest or usability may no longer be feasible due to the burden that preservation would place on the organization. This study will suggest fiscally reasonable approaches to preservation of objects of high value, passively retaining less-used objects in their original formats in the event they are needed, and deselecting those whose value no longer exists.

B. Definitions

To clarify the context in which terms are used:

- **Complex object**: a cohesive unit containing multiple parts that function interactively, use various format types (including but not limited to, text, image, audio, video and executables,) and may point to other objects. [1]

- **Digital object**: a string of bits that is viewed as an entity in its own right (e.g., a full-text document) though it may be a part of another digital object (e.g., an image that is part of a book), often with associated "metadata" and sometimes with terms and conditions (especially on access).[2]

- **Digital preservation**: Series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change.

    - **Long-term**: Continued access to digital materials, or at least the information contained in them, indefinitely.

    - **Medium-term**: Continued access to digital materials beyond changes in technology for a defined period of time, but not indefinitely.

    - **Short-term**: Access to digital materials either for a defined period of time while use is predicted but which does not extend beyond the foreseeable future and/or until it becomes inaccessible because of changes in technology.[3]

- **Emulation**: Preservation approach that involves the recreation of the technical environment required to view and use a digital collection. This approach uses software to mimic every type of application that has been written for every type of file format, making them capable of being run on whatever computing environment is current.[4]

- **Metadata**: Information which describes significant aspects of a resource. Used for resource discovery as well as to organize information required to successfully manage and preserve digital materials over time and to assist in ensuring essential contextual, historical and technical information are preserved along with the digital object.[5]

- **Migration**: Periodically moving files from one file encoding format to another that is useable in a more modern computing environment.[6]

- **Multimedia**: The combination of sound, graphics, animation and video. Multimedia is a subset of hypermedia, which combines these elements with hypertext.[7]

- **Preservation metadata**: a structured way of describing the preservation management requirements of digital resources. Such metadata might be used to store technical information that supports preservation decisions and action, to document preservation action taken such as migration or emulation, to record the effects of preservation strategies, to ensure the authenticity of digital resources over time, and to note information about collection management and the management of rights.[8]

- **Refreshment**: Periodically moving a file from one physical storage medium to another to avoid physical decay or obsolescence of that medium.[9]

C. Conditions Known to Exist

- DTIC has not yet devised an enterprise-wide policy on non-print preservation;
- this is a diverse collection whose physical condition is unknown;
- preservation research is being conducted world-wide, but at this time, best practices are subjective—what works for one organization does not necessarily fulfill another's needs; and
- DTIC online records are the customers' primary point of access to objects.

D. Assumptions

Certain assumptions were present at the beginning of analysis:

- media would not be useable due to physical deterioration or changes in operating systems and hardware;
- objects in the collection use a plethora of systems and software, some of which are not now commonly available;
- contributors and consumers expect DTIC guidance on preparing products for preservation and long-term access;
- bibliographic records—customers' point of access to collections--may not accurately reflect the product the customer will receive when s/he places an order; and
- the Defense Virtual Library (DVL) might be used for management of this collection.

E. Context and Scope

Sixty-six objects were examined. A cross-section was randomly selected, so that a range of unique attributes, including a variety of file formats, file

size, variety of content and level of complexity, could be observed. Since complete order records for non-print were available beginning in 1991, that was selected as the start date for statistics. The age of any media is often a factor in its stability, but it is not the only factor. It is conceivable that media received by DTIC at any time are potentially in danger of physical deterioration or technological obsolescence. This paper strives not to determine which individual objects are still functional; it attempts to address the larger issue of DTIC policy and how non-print objects could be perceived and managed.

Each object to be examined was opened and used as a customer might use it; a survey was completed for each object. (See survey questions at Attachments 1, 2 and 3.) Data was collected on file formats, total file size, presence of links and accompanying documentation, parent-child relationships between files, the enduring nature of the content, application software needed to open the files, and other attributes. On August 29, 2002, 2,337 records for non-print were found in WED.

Objects Contributed to the Collection, 1991-2001
(Composition of the collection and number of each format that must be dealt with.)

Media Type	Records Created or Updated, 1991-2001[1]	Replaced in the Collection[2]	% of Total Media Collection[3]	# Announce -ment Only[4]
CD-ROMs	691	42	32%	95
Diskettes, PC				
5.25", Hi Density	96	15	5%	9
5.25", Low Density	186	12	9%	15
3.5", Hi Density	757	76	34%	5
3.5", Low Density	46	1	2%	2
Diskettes, Mac				
3.5", High Density	29	0	2%	2
3.5", Low Density	19	0	1%	0
Magnetic tapes	132	59[5]	7%	18
Tape cartridge	1	0	1%	0
Videotapes (VHS)	148	1	7%	1

Data sources:
[1] DTIC History file

[2] DTIC ordering History file

[3] Based on collection of 2342 objects, record creation or update dates: DTIC History file, 1991-2001.

[4] WED (Distribution Code 21 and Media Code)

[5] Converted to tape cartridges in December, 2000.

Records created and updated between 1991 and 2001
(Indicates trends in contributions.)

Media Type	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01
CD-ROMs	7	6	15	59	26	19	29	46	24	169	298
Diskettes, PC											
5.25", Hi Den	20	22	32	17	3	3	0	0	0	0	0
5.25", Low Den	84	48	38	8	6	1	0	0	0	0	0
3.5", Hi Den	1	26	44	127	66	170	94	81	19	73	54
3.5", Low Den	12	13	9	4	6	1	0	0	0	1	0
Diskettes, Mac											
3.5", Hi Den	0	4	9	10	4	2	0	0	0	0	0
3.5", Low Den	7	4	5	2	1	0	0	0	0	0	0
Magnetic tapes	24	18	27	16	13	18	15	1	0	0	0
Tape cartridge	0	0	0	1	0	0	0	0	0	0	0
Videotapes (VHS)	25	17	16	16	11	15	16	8	3	4	13

Source: Record Creation Date in DTIC History file.

Distributed between 1991 and 2001
(Indicates trends in customer use of the collection.)

Media Type	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01
CD-ROMs	7	0	1701*	0	93	200	332	187	124	250	193
Diskettes, PC											
5.25", Hi Den	11	21	68	76	45	24	19	17	15	7	2
5.25", Low Den	131	716**	121	64	63	32	5	7	4	2	2
3.5", Hi Den	1	129	307	421	414	356	630	200	110	110	49
3.5", Low Den	26	117	140	34	17	14	7	7	1	2	2
Diskettes, Mac											
3.5", Hi Den	0	1	11	29	10	11	3	1	0	2	0
3.5", Low Den	2	28	24	14	8	6	0	1	0	0	0
Magnetic tapes	32	67	35	23	7	2	4	1	1	1	0
Tape cartridge		0	0	0	0	0	0	0	0	0	0
Videotapes (VHS)***	258	52	42	522	538	258	126	109	45	28	8

Data source: DTIC History file, 26 August 2002.

* Includes 1,697 copies of ADM200140: Corporate Information Management (CIM) Help Disk. (Replaced by ADM000444 per WED)

**Includes 598 copies of ADM000095: PCB Tester Selection for Future Systems/DUT/ATE Matching Algorithm (DAMA). Proprietary Version. (No longer available per STINET and WED.)

***Three titles account for 60% of all videos ordered:

ADM000509 Defense Technical Information Center: Fifty Years of Excellence in Information Service (Closed Caption) 115 copies

ADM000345 DTIC: Information Today--Success Tomorrow (Closed Caption) 383 copies

ADM000338 DROLS: An Introduction to Searching 691 copies

Security Classification of Objects
(Indicates the proportion of the collection that can be accessed via Unclassified means.)

Media Type	Unclassified	Confiden.	Secret	Total
CD-ROMs	750	3	3	756
Diskettes, PC				
5.25", Hi Den	62	0	2	64
5.25", Low Den	83	0	6	89
3.5", Hi Den	645	13	19	677
3.5", Low Den	31	0	0	31
Diskettes, Mac				
3.5", Hi Den	25	0	0	25
3.5", Low Den	10	0	0	10
Magnetic tapes	34	0	5	39
Tape cartridge	1	0	0	1
Videotapes (VHS)	125	0	1	126

Data source: WED

Ordering History

(Indicates relative customer interest in specific media. Possible that low interest levels indicate inability to use media, not low interest in subject matter.)

Format	**Average number of times distributed since 1991**
CD-ROMs	5[1]
Diskettes, PC	
5.25", High Density	3
5.25", Low Density	6[2]
3.5", High Density	3
3.5", Low Density	8
Diskettes, Mac	
3.5", High Density	2
3.5", Low Density	4
Magnetic tapes	1.5
Tape cartridge	0
Videotapes (VHS)	14[3]

Data source: Number of objects distributed 1991-2001, divided by the number of objects in the collection less "Announcement Only" citations. A few titles account for the majority of objects distributed. In the following three cases, the average times distributed was reduced to:

[1] 2.3 times if ADM000444, *Electronic College of Process Innovation (ECPI) Release 2.1: Achieving Breakthrough Improvement* is removed from the count.

[2] 3 times if ADM000095, *PCB Tester Selection for Future Systems/DUT/ATE Matching Algorithm (DAMA*) is removed from the count.

[3] 5.5 times if the following VHS tapes are removed from the count:

- ADM000509, *Defense Technical Information Center: Fifty Years of Excellence in Information Service*,
- ADM000345, DTIC: Information Today--Success Tomorrow, and
- ADM000338 *DROLS: An Introduction to Searching*

Part II: Findings

A. Physical Media

As expected, the collection is comprised of a wide array of media and file formats. Each object tested was useable and appeared to operate as designed; none appeared to have obsolesced to the point of full or partial failure. If the object opened without apparent problems, it was assumed to be operating as intended by its creator.

B. Objects

1. Due to variation in file formats, many different applications are required to open objects. The collection is comprised of objects that include but are not limited to:

- Files requiring use of proprietary software;
- Flat text files;
- Commercially available application formats such as Microsoft Office and AutoCad;
- PDF files displayable in various versions of Adobe;
- Still images in an array of image formats;
- Complex digital objects with a combination of file formats;
- Analog moving images (videotapes); and
- Digital moving images.

2. Metadata characteristics of these objects:

- Preservation-level metadata is not currently accepted by DTIC, and STINET records do not now accommodate it.
- Current metadata may not accurately represent the object, or the information may not always be relevant. For example, in some records the capacity of the media is cited but not the size of the file on the media.
- Application software needed to open an object is not always immediately available, either with the object itself, or available from other sources.

C. Internal Policy and Practices

- DTIC does not evaluate objects for preservation; once it is part of the collection, it is never de-selected. The collection contains objects that have dubious preservation value, such as a 1996 Defense Acquisition University course catalog and regional lists of maximum allowable charges in 1996 for CHAMPUS patients.

- Most of the collection is used very little. Fifty-six percent has been ordered three times or fewer, including 31% that have never been ordered.
- Contributions are accepted in a wide range of formats, which narrows the pool of customers able to use the object. Early guidance is not given for creating products that can be used by a wider audience. Current guidance makes it easy for contributors to prepare objects in a way that best suits their requirements, but that does not necessarily make the product easy for the majority of customers to use.
- In some cases, external software or file viewers are required to use some objects. These may currently be available for free over the Web, but there is no guarantee they will remain either free or available.
- In some cases, the same object is available from DTIC in multiple file formats. For example, an object might be accessible online in PDF, on CD and on diskette.
- Records for objects available in multiple formats are not always consistent. ADM000847 (CD-ROM) is also available in text as ADA353650. However, the record for the paper version does not reference the CD-ROM version; other metadata—including title—is not identical, which can be confusing to customers.
- Customers may not realize additional documentation is available as some records are not cross-referenced. An example is the record for ADA324292 which does not reference diskette ADM200456 (Export Controlled). Unless customers know to check, they may not realize that the diskette exists until they read the full document. Customers must then place (and wait for) a second order. Many records include this information, but not all.
- Records may indicate a delivery format that most customers could not use (i.e., 5¼" diskettes) when, in fact, delivery is actually on more modern media. As each object is ordered, the request is fulfilled with the more current media (i.e.; 3 1/2" diskettes.) But there is no indication to customers that newer media is being shipped until it arrives in the mail.

D. Customer Access to Information

- Customers do not have the option of ordering an object in the media format that best suits their need (assuming that the media is appropriate to the object) or downloading it. For example, it is not appropriate to transfer a 30 moving image file to diskettes. Customers are currently able to acquire objects only in their native formats--those in which they were originally contributed.

- DTIC policy does not cover retention of the original's functionality. True preservation protects all aspects of a digital object from deterioration, including content appearance and functionality. Emphases and resource allocation must be decided before a program decision can be made. Is it more important to devote all resources to retention of the functionality of a few originals? Or is it more important to do as much as possible with the greatest number of objects, giving customers the widest access? If the later is chosen, chances are reduced that the object will be available indefinitely.

E: Challenges

- It is costly to preserve, not only at the beginning of a program but over time, because preservation must repeatedly be re-accomplished. Greater accessibility and usability are by-products of preservation, but achieving that will consume more resources than maintaining a stand-alone collection. There is no guarantee that resources to support such a program will be plentiful over time. The organization must make a value judgment, on the collection as a whole, or on a title-by-title basis to determine whether access over the long-term offsets the cost of achieving greater accessibility.

- DTIC policy of accepting contributions in diverse file formats increases the cost of management. Since preservation of each object requires individual attention, the wide array of formats can impact resources significantly.

- There is no single best approach to preservation; different approaches are used for different needs. Because the non-print collection is broad, no single approach can be used in all cases. The best solution appears to be a mix of approaches. The three primary approaches are briefly described:

- *Refreshment* moves a digital object from one medium to another with no change in the bitstream to avoid physical decay or obsolescence of the original medium. Obsolescence is continuous, so media refreshment will be necessary throughout the object's lifecycle.

- *Migration* makes existing applications and data work on different computers or operating systems by moving files periodically from one encoding format to another, making the object useable in each succeeding computing environment. For example, a spreadsheet in an old version of Excel might be moved to a newer version on a newer operating system. Migration reduces the problem of files encoded in a wide variety of older formats by gradually transitioning them all into a limited number of contemporary formats. This is currently the favored approach, but it does have specific dangers; namely data loss and loss of original functionality (the look and feel of the original.)

- *Emulation* seeks to solve some of the same problems as migration, but by focusing on application software rather than files containing information.[10] A major issue in digital preservation is determining what aspects of an object besides intellectual content needs to be preserved. The "look and feel' of an object may be important, as may its interactivity. Leaders in the preservation field fear that migration can destroy these aspects; that emulation is the only strategy guaranteed to preserve them. This problem intensifies as more complex digital objects (multimedia and hypermedia) are created.

  There are two schools of emulation thought. One approach advocates construction of software that mimics computer systems and applications, recreating the way that each file format runs on each new computer system. In effect, this recreates the original computing environment for each new generation of computers.

  IBM is exploring another emulation process for the Koninklijke Bibliotheek (National Library of the Netherlands) and the British Library. IBM is attempting to develop a Universal Virtual Computer (UVC) to emulate programs. The manufacturer of each application or platform provides an emulator of its current application or platform in UVC code. This code emulates the device on the application or platform

that controls the software or equipment. As new machines are built, or new applications are written, their manufacturers will produce UVC interpreters for later use in emulation processes.[11]

Worth watching is the Creative Archiving at Michigan and Leeds: Emulating the Old on the New (CAMiLEON) project, a joint National Science Foundation (NSF)/Joint Information Systems Committee (JISC) project to explore different emulation strategies. This project seeks to put emulation theories into practice.[12] It has not yet produced tangible results, but expects to:

- evaluate publicly available emulators,
- explore emulator development,
- conduct test cases from both technical and user perspectives,
- conduct trials comparing original systems with emulated versions of those systems, and
- undertake cost-benefit analysis of emulation vs. other digital preservation strategies.[13]

Both emulation approaches assume that one solution will work in all cases. And emulation may eventually develop that way. But until that happens, conventional wisdom is that a combination of refreshment, migration and emulation strategies will be used.

F: Cost

The emphasis of digital preservation is access over time; it involves continued attention to an object until the object's intellectual content is no longer needed. Establishing the initial technical infrastructure for preservation and building a digital collection can be costly.[14] Precise costs cannot be determined, partly because it is difficult to isolate the precise part of an agency's budget that covers preservation. "…all the aspects and tasks involved in digital collection management are closely interlinked, making it very difficult to identify those elements which relate solely to preservation."[15] Costs depend on many factors, including the objects selected for preservation, determination of appropriate strategies, validation of the completeness of the object as it enters the repository, production of metadata, digital storage and administration.[16] Additional costs are incurred in training personnel who will be dedicated to care of the collection; and in maintaining equipment and networks.

Forecasting the cost of a project is a subjective exercise. Because each project is different; using data gathered from other projects is not likely to be an accurate measure of the project at hand. Perhaps the best way to determine an accurate cost is to create a representative sample of objects selected for preservation.[17]

Rapid technological turnover further increases costs. To be effective, the most current standard must be maintained. Expectations for access will increase as users continually increase their technological sophistication. Digital preservation is a relatively new science, long-term cost models do not exist. Assumptions that can be made:

- Technology will change often throughout the foreseeable future.
- Use of uniform formats and encouragement of best practices in data creation and preservation preparation is a major factor. Not only do uniform practices facilitate access, they also make a collection easier to manage and present fewer problems in later interoperability. Use of standards creates a greater chance of achieving economy of scale.[18]

While preservation may be costly, the price of not preserving can be just as high. Much research cannot be reproduced; when that documented knowledge is lost, it may not be recoverable. Common sense says that preservation is less costly today than it will be in the future. In addition, having information in digital form will become more and more important as time goes on. Retention of the full purpose, functionality, look and feel of a preserved object will be a critical support to customers. Action taken now will only benefit DTIC by developing knowledge necessary to preservation of future work.

G. Other Issues

- DTIC customers prefer products that work easily in their systems, and from which they can pull non-copyright protected information to insert into other products. They want--and often, can only afford to maintain-- products “just in time” not “just in case.”
- Because it requires constant examination, updating and reformatting, preservation competes with other resource requirements, not only when the object is created but throughout its life-cycle.
- A small sample of the collection was evaluated. All media in the sample were in good physical shape and were useable. It is possible that some non-tested objects may have deteriorated beyond usability, but it would be necessary to test the entire collection to make that determination.

Section II: RECOMMENDATIONS

A. Introduction

The Section recommends actions to preserve long-term customer access to information in the DTIC non-print collection. Some general observations and recommendations are given in this part; specific actions are suggested in parts B and C. DTIC's print collection is fundamentally different from the non-print collection in size, complexity, potential for loss and resources required to maintain it over the long-term. These recommendations may require a policy change for the non-print collection. These changes may be applicable but are not suggested for the print collection for a number of reason, but primarily because at this time, there is much less urgency to address preservation of print/text than non-print.

- A "magic bullet" preservation strategy does not exist.[19] Multiple strategies reflecting organizational needs must be identified and employed for this collection.
- Solutions…evolve over time.[20] To avoid altering entire systems during this evolution, modular interoperable solutions should be used in design of ingest systems, metadata management, and user interfaces. This not only permits relatively fast and cheap modifications and upgrades, but it also allows implementation of new technology without complete reworking of old systems.
- Digital objects never rest. Every preserved object must be revisited periodically for update. All objects, regardless of format, will eventually obsolesce.
- The best time to compile preservation metadata is at the time and place of object creation, when file characteristics can be recorded more easily.
- Preservation enhances interoperability. A number of domestic and international groups are struggling to determine standards and direction for digital object preservation. Commitment to participation in the larger community will ultimately pay off through easier access, use and sharing.
- The subject matter of all objects is not created equal. Given the concerns of preservation, each object must be evaluated to determine whether active preservation, passive retention or deselection is merited.

B. Short-Term Recommendations (Six months)

1. Evaluate the collection item by item to determine suitability for retention.

To advance technological innovation, DTIC encourages the defense community to share its scientific, technical and analytical information by contributing appropriate materials to DTIC. In addition, DTIC supports the DoD Acquisition Program[21], which necessitates collection of information about Defense planning and budget execution. This broad role encourages a wide range of contributions including some that are tangential to DTIC's mission, such as a 1995 Federal acquisition course catalog (ADM000464), lists of FY96 DoD prime contractors sorted in various ways (ADM000500 et al.) and a credit card training course (ADM000820.)

In addition, almost one third (31%) of the collection has never been ordered. More than half has been ordered three times or less. Some objects require operating systems and application software that are unavailable to customers; it is not likely access to these systems will surface in the future. Maintenance of a large collection of contributions created in such a diverse range of formats and media make it difficult and expensive to preserve all of it.

A. Determine criteria for evaluation.

To determine if objects are worth the resources required for preservation, an object by object assessment using the following criteria will determine which of the current collection is suitable for preservation; or at least minimal attention:

1. Is the media in good physical condition? Is it accessible on currently available platforms? Is it likely that this hardware will be continuously available; or is the hardware becoming less common, as in the case of VHS tape players being replaced by DVD players? (Video vendors are removing VHS tapes to make room for DVDs. It is not unlikely that DoD will adopt this trend as well.) Is it likely that this media will be compatible with future hardware or systems as in the case of CD-ROMS accessible on DVD drives?

2. Does the object have external (i.e.: Web) links, or other functionality that must be retained in order to use the content as intended? If these links or functionality are already broken or cannot be preserved, how will that impact future use?

3. Is application software, needed to use this object, still commonly available? Does DTIC have access to this software? Do customers have access?

4. What bearing does the content have on DTIC's Defense RDT&E mission? Does it have continuing technical, research or scientific value? What current or potential impact does it have on DTIC's collection (i.e.: how will DoD be impacted if DTIC does not own this?) Is it likely to be needed at a later date, either as a historical record or to illustrate a current technology to future users? Or, is the content ephemeral with little continuing value? Is it relevant only to the time period in which it was created? Was it created for an audience that no longer exists, such as a course catalog? If that audience no longer exists, is it likely that the content has been archived by the originating agency? Has the content been superseded, corrected or updated by a later edition?

5. Does an agency other than DTIC have responsibility to preserve the content? What value does the originating agency place on the object? Does it consider the object worthwhile enough to preserve? If the agency cannot be contacted, does DTIC feel the content's value:

a. requires future action, or

b. requires retention in its original format without further action?

If the agency cannot be contacted and there is no evidence that the object has any continuing value, does DTIC have the drawer space to retain it into perpetuity?

6. Is this object copyright protected? If so, is it feasible to contact the owner for permissions or changes?

7. Has it been ordered from DTIC at least once in the preceding five years?

8. What input can the contributor furnish? Does the contributor believe the information has continuing value? Have corrections been made to the content? Does the contributor now distribute the same content on more current media?

DTIC should consider de-selecting objects that do not fit minimum criteria. Criteria are not absolute; objects should be weighted subjectively so that each object is given a separate evaluation based on its suitability for retention and content value to DTIC's mission. Once the initial review is completed, a regular (perhaps biennial) review schedule should be instituted to examine all objects to determine whether they have become

candidates for further action. Future selections should be examined carefully for their applicability to the collection policy. Decisions should be documented in the online record. In the event that the object is withdrawn, records should be retained in the database with the note: "DTIC does not have this document."

2. Evaluate each object to determine how it should be retained.

Each object must be individually evaluated to determine whether it will be retained in its original format and furnished to customers in only that format; made available through STINET; or held on an in-house server and copied to the media of customer choice, assuming the object is suitable for that media. To determine the group to which an object belongs, the following criteria can be applied:

A. Can the entire object be made available in STINET given its classification and distribution limitations?

B. Does the content appear to merit immediate customer access? Does the "Acquisition Reform Roadshow" (ADM200524) bear the immediacy of the papers presented at a conference titled: "Minimizing Chemical Warfare Threats Through Development of Advanced Medical Countermeasures" (ADM001167?)

C. Is the object size or length suitable to online display? For example, the average length of a videotape accession in the collection is 37 minutes, with lengths varying from three minutes to almost 13 hours. At the 2002 Annual Users conference, customers commented that even the limited video in DTIC's STINFO Distance Learning course required so much time to buffer and stream that they avoided viewing it. Making lengthy or complex objects available online is not a practical solution. In this case, a more suitable approach is to provide an excerpt or thumbnail summarizing the content. Customers might then have the option of ordering the entire object on the removable media they prefer.

D. What resources are needed to make the object accessible through STINET? Content subject matter, customer order history and likelihood of future customer interest are factors in determining whether the customer's need balances the resources required. Is it more worthwhile to have it accessible online than to have it available only in its original format? For example, just the cost of digitizing an (analog) VHS videotape in RealMedia for streaming and MPEG 1 for preservation may well be more than its anticipated potential value to customers:

ADM000725: *Reinvention Revolution Conference Address and Town Hall Meeting with Vice-President Al Gore* (73 minutes, 1997)
Estimated cost to digitize and process: $1,307.00
Number distributed since acquired by DTIC: 3

ADM000069*: New Leader: a Documentary on the Orleans ROTC Battalion* (30 minutes, 1991.)
Estimated cost to digitize and process: $537.00
Number distributed since acquired by DTIC: 1

E. Which copyright or intellectual property concerns must be addressed before the content can be used in any form other than the original?

F. Does the contributor have any restrictions on how the content will be retained?

As stated before, the long-term value of content varies. Limited resources will require value judgements on what level of access can be provided to customers.

3. Enhance and expand metadata for preservation and access.

True preservation requires that descriptive, access and preservation metadata be associated with an object. Metadata provides details of the object's content, characteristics and provenance, remaining with the object throughout its life-cycle. Metadata affects users' ability to access it over time. For example, it is important to know which version of application software a document was prepared in, in order to use all its functions.

Older DTIC metadata could be corrected, expanded and enhanced. As the non-print collection has become more technically sophisticated, metadata for the non-print collection has become much more relevant and specific. Much of DTIC's earlier metadata is accurate but incomplete or irrelevant. For instance, diskette capacity may be given but not the object's file size. Although great care is taken to make sure new records contain critical information, some older metadata is outdated. For example, records for 5¼" diskettes do not indicate to the customer that they will be shipped on 3½" diskettes. Records for magnetic tapes converted from nine track open reels to cartridges in 2001 still indicate the format as open reel tape. Uninitiated customers will believe that, if they do not own the applicable equipment, they are unable to use the object. To prevent confusion, metadata should be included in records to indicate exactly what the customer will receive.

Several technical metadata schema have been developed by members of the international preservation community, including one developed for the DVL project. It outlines initial metadata needed for long-term preservation.[22] DTIC could begin to build basic preservation metadata by revising the contribution process and Form 530 so that contributors are prompted to provide preservation level metadata. Alterations to EDMS so that it accepts and displays additional metadata are complex and expensive. Realistically, resources may not be available for programming changes, but assuming EDMS use continues, it may be more productive in the long run to make the changes as soon as possible.

Technical metadata should be displayed to customers performing STINET searches. Metadata elements including *Context Relation Manifestation* (describes related material, both digital and analog) and *Provenance Action Method* (details of any process applied to a digital object's media or file, including media refreshment or migration) will keep customers informed about whether they have the technology to use the object and will help them make value judgements regarding content.

Assuming that current trends continue, DTIC will eventually manage content in XML. The technical capabilities of contributors can be exploited by accepting metadata in eXtensible Markup Language (XML). Aids such as automated input forms like the SF-298 can be developed to assist customers in supplying XML-formatted data. Once XML metadata is compiled, it is relatively easy to transfer into an array of desired formats, enhancing interoperability. DTIC is developing document type definitions (DTDs) based on World Wide Web Consortium (W3C) recommendations. These will allow easy manipulation and mapping of XML-based metadata to a variety of schemes, including COSATI. A working draft of a Technical Reports DTD, based on the DROLS full citation (1F) display format, is now in coordination. As technology changes, the number and format of required metadata elements will evolve. The flexibility built into XML makes it easy to adapt to that change.

4. Make more extensive content available online.

Content received by DTIC on paper is scanned, put into TIFF and PDF, and made available online as well as output to paper and microfiche. But at this time, DTIC disseminates content received on removable media in only one format—its original.

Objects in the non-print collection selected for long-term retention would become media independent if they are transferred to a server, which would permit wider customer access. It would no longer be necessary to maintain specialized hardware, and objects could be output to customer-defined media that is suitable to their size and file format. Customers searching

STINET could link from the record to the server version of the object, just as they now link to the PDF file of a text document. The server version might be the full object, a clip, a thumbnail or other excerpt, depending on object size. Objects retained on a server reduce the number of possible formats that must be managed, which in turn, lightens the resource burden.

DTIC owns or is building three Rimages, the largest with 3-18GB SCSI drives, to hold content acquired on diskette and CD-ROM. The memory on these machines is capable of significant expansion. Current practice is to load selected digital objects to these drives, purging less critical content to make room for additions. Customer orders for these objects are filled by copying from the Rimage drive. Criteria for determining which objects to load or purge are currency of the object, customer demand, and available disk space. Master copies of objects are retained in their original format. Maintaining objects on servers and networking those servers would make digital objects more accessible.

A customer could also select a copy on preferred media from the Document Order page. By making an object media-independent, it becomes much easier to reproduce an object on media other than the original. By permitting a customer to select the media on which a requested object is distributed, DTIC opens content to a wider audience.

If DTIC is to appeal to its customers, it must make information available in the forms they want and need. Since customers need a collection that is broad as well as deep, it follows that it is critical to increase the number of items accessible by customers. Statistics show that accesses of electronic resources far outnumber purchases of hard copy. The Web has become the world's communication vehicle of choice. It is possible that non-print objects not currently being ordered might be used more if they were accessible online or in a customer selected format.

5. De-select identical content in multiple forms.

In some cases, objects with identical content were contributed to DTIC in multiple media formats, presumably to give customers flexibility in which format the information could be delivered. But if Recommendation 4 above is put in place, this will no longer be an issue. The most advantageous instance of an object should be retained in the collection, with extra instances de-selected and their records withdrawn or annotated as appropriate.

6. Insure that adequate training and tools are available to collection managers.

The most recent practice for collection management is to divide responsibility for cataloging between the collection manager and a DTIC cataloger. The collection manager is responsible for technical metadata elements, while the cataloger describes the intellectual content. Historically, non-print collection metadata has been formatted in COSATI which is not as rigorous as other cataloging schemes, namely MAchine Readable Cataloging (MARC.) If DVL and its MARC-based bibliographic control will be used to deliver multimedia content, the skill sets of either the collection manager and/or cataloger(s) must include MARC cataloging skills.

Shared cataloging of objects in the collection is unlikely to be available outside of DTIC, so customers will expect to be able to download quality bibliographic control records from DTIC for their own systems. The non-print manager should have at least rudimentary training in principles of information organization and bibliographic control including application of subject headings, thesauri, classification schemes, cataloging standards and bibliographic utilities. Such training is available from the University of Maryland College of Information Studies, or Catholic University of America's department of Library and Information Science. Software tools might also be used to assist construction of quality records. Minaret (www.minaretsoftware.com,) which converts plain text to MARC records, is already owned by DTIC; MARC Magician by MITINET (www.mitinet.com) could be purchased for approximately $1,100. Other tools assist creation of eXtensible Markup Language (XML) records.

C. Long-Term Recommendations

1. Determine customers' needs.

An in-depth understanding of customer requirements will help determine how the non-print collection should be developed and will aid planning of a preservation program. Should DTIC appeal to the bench technician, the scientist or the librarian? How about the soldier or marine in the field? The scientific community, the intelligence community and academia are customers, but what are their needs? What are the needs of Defense contractors and the general public? How are contractors defined? What security issues for transmitting non-publicly available information have arisen since September, 2001?

Although customers, as represented by members of the Users Council, unanimously support preservation on a conceptual level, they are concerned less about preservation's long-term benefits than its shorter term payoff—wider, more efficient access to information. They would like to have formats like still and moving images delivered online for re-use in their own products.

Better knowledge of customers will also result in better knowledge of the contributions those customers can make to the collection.

2. Enhance contributor guidelines.

The more uncommon the file format, form of media or required application software, the more cost and labor-intensive it will be to preserve an object. Rare formats or unfamiliar applications increase the likelihood that customers will be unable to use an object. The traditional practice of accepting all contributions may open the door to many contributions, but today's proliferation of technology increases the level of resources needed to manage and preserve such a variety of formats. As staffing levels fall, the balance between resources and need becomes even more pronounced.

Giving contributors a better concept of preferred formats--of what will more likely stay useable into the future--should increase the likelihood that DTIC will be able to maintain those objects into the future. By requesting that contributors adhere to a relatively narrow set of criteria when preparing file formats, media formats and metadata, DTIC is likely to decrease its staffing and preservation costs. These criteria could be developed by a cross-Directorate team and include input from selected contributors. Providing these criteria at the time digital objects are created rather than at the time they arrive at DTIC makes the process more efficient and cost-effective.

Not all contributors will be able to conform to DTIC guidelines, which can make their products unattractive to consumers. Creators may be required to produce in agency-dictated formats; products may require antiquated software or operating systems. These creators would still be able to contribute, but they would have to understand that their objects may become inaccessible over time.

3. Develop more narrow content selection criteria.

DTIC's content selection criteria are broad; policy is to permanently retain all contributions that can generally be defined as scientific and technical information. Current policy focuses mainly on physical characteristics, object type (i.e.; annual reports, lessons learned, DoD-related training courses, directories and indexes) as well as subjective qualities like intended use, usefulness, extent of activity and precedent. While these are valid selection criteria, a deeper approach to selection would also address the object's long-term value to customers.

Although not an archive, DTIC is considered an ultimate resource, able to furnish information that cannot be found anywhere else. But the cost of maintaining a small number of non-print objects without weighing their content value is expected to be high, especially when each requires special handling. Collection review is time consuming and difficult, but as technology evolves, review will become increasingly critical.

Current content selection criteria (www.dtic.mil/dtic/pubs/contribute) could be expanded to include:

1. Demonstrated relevance to DTIC's RDT&E mission.
2. Relevance to missions of DTIC's customers.
3. Length of time the content is likely to be relevant to users.
4. Likelihood that the information will be available to DTIC customers from another source on a long-term basis.

4. Devise an overall preservation policy.

The longest-term issue is to roll all of the above into a formal preservation plan that covers all aspects of the long and short-term needs of DTIC and its contributors and customers, balanced against a reasonable expectation of resources, desired quality and availability of collection and services, and coordination of DTIC actions with the larger community.

5. Watch the technology.

Preservation is far from an established science. Until it is more refined, DTIC should continue to monitor the literature and track changes in technology and schools of thought. Monitoring prevailing trends will help DTIC plan a collection that is preservation-ready. Plotting preservation

goals and objectives will help to insure that resources and collection are available to customers over the long-term.

SUMMARY

It is difficult to separate DTIC's customer support from its approach to preservation. The way that DTIC addresses preservation impacts the information it ingests, the way it manages that information and how it disseminates it to customers.

An editorial in the May, 2002, issue of D-Lib magazine synopsizes the direction that DTIC's efforts should take: preservation is "extremely complex and presents enormous problems, but it is important to make a start."[23] The earlier the organization makes its start, the less difficult it will be to manage preservation down the line.

A specific definition of preservation as it applies to DTIC is needed before a course of action can be prescribed. What is the Center's goal? Is it to preserve resources so that future researchers can use them–displaying, searching, browsing, manipulating them as their creator intended using the same processes with which they were created, as they can do now with the original object? Or should it be a more modest approach, focusing only on preserving as much as possible of the *knowledge* contained within the object, preserving the object's description (metadata) and keeping the essential functionality intended by the creator? The first step in any preservation program is determining an approach.[24]

True preservation, that is making digital objects continually and fully accessible over the long-term, appearing and functioning as they were originally intended, will require a long-term, resource-heavy organizational commitment. It will not be an insignificant effort. Some of the current collection is relatively simple and would require little effort to preserve as creators intended (i.e. flat text/ASCII files on 3½" diskette), but the larger part of the collection consists of more complex objects, for which preservation strategies like emulation or migration will be required.

So, what to do? The bottom line on preservation of this collection is that sustaining future access through true preservation will be expensive, complex and never-ending. Re-formatting media, digitizing content, acquiring equipment and re-programming existing systems to accommodate additional metadata are just a few steps toward accomplishing this. And the reality is that, with future resources likely to be scarce, full preservation of the entire retrospective collection is not a prudent choice. Addressing preservation through future contributions is more likely to yield a greater return. And the sooner that this can be addressed, the more economical it will be. If retrospective preservation of

any kind is to be accomplished, it is critical that action be taken soon on aging resources, even if that action is minimal, i.e. refreshing the media. For future contributions, the most effective step is to influence creation and contribution. This presents the opportunity to begin collecting content that is more likely to be sustainable over time; content that is more easily ingested, managed, preserved and accessed by customers. By limiting the variations in object file formats and media, and by dealing with preservation at the point of creation, the resource cost is borne in part by the contributor. Ingestion of preservation-ready objects now means that DTIC will not need to re-compile critical information later. DTIC will also have fewer issues to address later (outdated media, compiling metadata, formatting files, digitizing) when the cost per object has risen.

Given the uncertain science of preservation, DTIC has done what it can to prepare its Technical Reports collection for preservation. Use of standard file formats, use of microfiche as a preservation format and identification of preservation metadata elements are all positive steps. DTIC staff are developing expertise in preservation and are participating in Community activities that promote standards. But in terms of its non-print holdings, it is recommended that DTIC look forward rather than retrospectively. DTIC should build on its strengths, work to develop a collection that has the capacity to be preserved and develop new ways to deliver customer support, making it easier for customers to use its services.

REFERENCES

[1]Metadata Guidelines for Complex Digital Objects. Defense Technical Information Center, Defense Virtual Library. Silver Image Management, 2001, p. 3. Retrieved October 21, 2002, from Defense Technical Information Center Web site: http://dvl.dtic.mil/metadata_guidelines/dvl_cdo_guidelines19Jun01.pdf.

[2]Baeza-Yates, Ricardo and Berthier, Ribeiro-Neto. (1999). *Modern Information Retrieval.* Glossary. Retrieved June 5, 2002, from University of California, Berkley, School of Information Management and Systems Web site: www.sims.berkeley.edu/~hearst/irbook/glossary.html#D.

[3]Jones, Maggie and Beagrie, Neil. Preservation Management of Digital Materials. London: British Library, 2001, p. 10.

[4]Kenney, Anne R. and Rieger, Oya Y. Moving Theory Into Practice. Mountain View, CA: Research Libraries Group, 2000, p. 148.

[5]Jones, Maggie and Beagrie, Neil. Preservation Management of Digital Materials. London: British Library, 2001, p. 11.

[6]Sitts, Maxine, editor. Handbook for Digital Projects: a Management Tool for Preservation and Access, Andover, MA: Northeast Document Conservation Center, 2000, p. 160.

[7]Microsoft computer dictionary. Redmond, Washington: Microsoft Press, 1999, p. 301.

[8]Preserving Access to Digital Information: Preservation Metadata. National Library of Australia. Retrieved October 21, 2002, from: http://www.nla.gov.au/padi/topics/32.html.

[9]Sitts, Maxine, editor. Handbook for Digital Projects: a Management Tool for Preservation and Access, Andover, MA: Northeast Document Conservation Center, 2000, p. 160.

[10]Sitts, Maxine, editor. Handbook for Digital Projects: a Management Tool for Preservation and Access, Andover, MA: Northeast Document Conservation Center, 2000, p. 160.

[11]IBM Research: Long-term Preservation Project. (n.d.) Retrieved June 5, 2002, from http://www.kb.nl/kb/resources/frameset_kb.html?/kb/ict/dea/ltp/jointstudy/page4.html.

[12]*CAMiLEON: Creating Creative Archiving at Michigan and Leeds: Emulating the Old on the New.* (n.d.). Retrieved June 5, 2002, from http://www.si.umich.edu/CAMILEON.

[13]Granger, Stewart. (October, 2000). Emulation as a Digital Preservation Strategy. *D-Lib Magazine, 6, 10*. Retrieved June 5, 2002, from http://www.dlib.org/dlib/october00/granger/10granger.html.

[14]Russell, Kelly and Weinberger, Ellis. (May 31, 2000). *Cost Elements of Digital Preservation*. Draft. Retrieved June 5, 2002, from Curl Exemplars in Digital ARchives (CEDARS) Web site: http://www.leeds.ac.uk/cedars/documents/CIW01r.html.

[15]*Digital Culture: Maximising the Nation's Investment*. British Library: National Preservation Office, 1999. Retrieved September 25, 2002, from http://www.ukoln.ac.uk/services/elib/papers/other/jisc-npo-dig/cover.pdf.

[16]Granger, Stewart, Russell, Kelly and Weinberger, Ellis. (n.d.) *Archive Models, Cost Elements and Strategies for Digital Preservation*. Retrieved June 5, 2002, from http://www.si.umich.edu/CAMILEON/ Emulation%20papers%20and%20publications/ Cost_Elements.doc.

[17]Sitts, Maxine, editor. Handbook for Digital Projects: a Management Tool for Preservation and Access, Andover, MA: Northeast Document Conservation Center, 2000, p. 28.

[18]Jones, Maggie and Beagrie, Neil. Preservation Management of Digital Materials. London: British Library, 2001, p. 57.

[19]LeFurgy, William G. (May, 2002). Levels of Service for Digital Repositories. *D-Lib Magazine, 8, 5.* Retrieved June 5, 2002, from http://www.dlib.org/dlib/may02/lefurgy/05lefurgy.html.

[20]Wilson, Bonita. (May, 2002). Digital Preservation – A Long Journey. Editorial. *D-Lib Magazine, 8, 5.* Retrieved June 5, 2002, from http://www.dlib.org/dlib/may02/05editorial.html

[21]*What is Suitable for the Technical Reports Database?* Defense Technical Information Center. Retrieved October 21, 2002, from: http://www.dtic.mil/dtic/submitting/what_submit.html.

[22]Defense Virtual Library Project. *Technical Metadata for the Long-Term Management of Digital Materials: Preliminary Guidelines*. Silver Image Management, March 21, 2002. Retrieved September 25, 2002, from: http://dvl.dtic.mil/metadata_guidelines/TechMetadata_26Mar02_1400.pdf

[23]Wilson, Bonita. (May, 2002). Digital Preservation – A Long Journey. Editorial. *D-Lib Magazine, 8, 5.* Retrieved June 5, 2002, from http://www.dlib.org/dlib/may02/05editorial.html.

[24]Bide, M., Potter, E. J., and Watkinson, A. (September, 1999) Digital Preservation: an Introduction to the Standards Issues Surrounding the Deposit of Non-Print Publications: a Study Conducted by Book Industry Communication on Behalf of the British National Bibliography Research Fund. Retrieved on June 5, 2002, from Book Industry Communication Web site: http://www.bic.org.uk/digpres.doc.

Attachment 1: Diskette Inventory Checklist

Software and version in which it was created
Is executable software included as part of the object?
Minimum system requirements to open the file
Does it include an accompanying paper document?
Is this document required to operate the media?
AD# for other media shelved with this
Is there a paper equivalent of this object?
Does it contain internal or external links?
Unusual technical requirements?
File size of full object
Is viewing software included on disc?
Supporting software required: version #?
Are free viewers available to run it?
Equipment/platform requirements/specs:
Hard drive
RAM
Monitor, printer, sound/video card or other equipment requirements?
Is existing metadata record complete, accurate and updated to latest condition of object?
Any special requirements not included above?
How many copies have been distributed?

Attachment 2: CD-ROM Inventory Checklist

Software and versions in which it was created
Is executable software included as part of the object?
Minimum system requirements to open the file
Complete object size
Special instructions or commands needed to open or run the object
Number of files on each piece of media
Is accompanying document required to operate the media?
Which software is required to open and use the object
Is this software included on the media?
Equipment/platform requirements/specs:
Platform/Operating System
RAM
Hard Drive
Monitor, printer, sound/video card or other equipment requirements?
Does it contain any internal or external (Web) links?
Is existing metadata record accurate, complete?
Are viewers publicly available to run it? If not, who is eligible to acquire?
Are viewing or operating software included on disc?
Any special requirements not mentioned above?
How many copies have been distributed?

Attachment 3: Videotape Inventory Checklist

Media format of related pieces?
Are related pieces required to use the tape?
Media format
Length of recorded moving image on each tape
Does it include title slates?
Are there creation/production credits?
Is it available in any other format in the collection?
Is the image closed captioned?
Does it include audio?
Total length of tape
Lengths of other tapes in this set
Does the content stand alone? Or must other pieces be available for one to work?
Is special equipment required to run or view the tape?
Any special requirements not mentioned above?

Appendices

1. PERSONS CONSULTED *(Partial list)*

Nutan Carr, DLA Video Productions

Jack Jones, DTIC-E
Dolores Knight, DTIC-E
Claire Duong, DTIC-OCA
Margaret Clifton, DTIC-OMI
Cedric Hairston, DTIC-OMP
Peter Mount , DTIC-ZS
Mike Paige, DTIC-ZS
Jeff Davidson, DTIC-ZW

Connie McEowen, DTIC Users Council
Carol Ramkey, DTIC Users Council
Jeanne Rosser, DTIC Users Council
Karen Schaffer, DTIC Users Council
Robert Seidel, DTIC Users Council
Dan Sell, DTIC Users Council
Susan Tarbell, DTIC Users Council

2. Data Compilation

File formats, software and operating systems used by objects (as annotated in metadata) are listed below. The list gives some idea of the collection's diversity.

A. DISKETTE

SOFTWARE AND FILE TYPES IN COLLECTION

Acrobat 3.0
Acrobat Reader
ADA
Adobe Acrobat
AuthorWare 3.01 for Windows
AutoCad 12
Borland C++ source code compiler
Borland 4.0
Boxer
C
C Compiler
C++

C++ compiler
Clipper 5.2 D
Clipper/C
Comet—Fortran computer program simulation programs available from EPA web site via AFCEE/ERT home page
DAT data files ASCII
DBase III+
DBase IV 2.0
DBase 4.1 runtime
Design IDEF 3.1
ECON Pak for Windows
ERWin 2.1
Fortran
Fortran 77
Fortran compiled with MS Fortran power station, 4.0
FoxPro Dbase
GeoCoPS
HTML
HTML 3.0
Hydroform
Ithink Analyst 4
Lotus Approach
MATLAB 4.2+
MATLAB 5.0
MS Access (version unknown)
MS Access 2.0
MS Access 7.0+
MS Excel (version unknown)
MS Excel 4.0+
MS Excel 5.0
MS Excel 7.0
MS Excel 98
MS Powerpoint (version unknown)
MS Powerpoint 4.0
MS Powerpoint 7.0
MS Powerstation compiler
MS Word (version unknown)
MS Word zipped
MS Word unzipped
MS Word 2.0+
MS Word 2.1
MS Word 3.1
MS Word 6.0
MS Word 7.0

MS Word 95
MS Word 97
MS Word 7 for Windows 95
NCSA Mosaic
NDP Microway
Object Vision 2.1
Online Hypertext System 1.0
Pagemaker 6.0
Pascal Spreadsheet
PDF
PKzip
PKunzip 2.04
Postscript
Printfile in HCL 4
“proprietary software application”
Quick Basic 4.5
System Architect 3.X+
Sound Designer
SQL Base
TAC 3
TMA software compiled in MS Fortran 5.1
Vega or better graphics
Visual Basic 3.0
Windows DDE
Windows Installer 1.1
WinZip
Wordperfect (version unknown)
Wordperfect 5.0
Wordperfect 5.1
Wordperfect 5.2
Wordperfect 6.0
Wordperfect 6.1
Wordperfect Envoy runtime Viewer 1.0
Zip2EXE self-extractor creator
Wordperfect

OPERATING SYSTEMS USED BY OBJECTS IN COLLECTION

AIX
Cray compatible workstations to run OSC-TPV Cray
DOS 3.0+ not Windows
DOS 3.0
DOS 3.31
DOS 5.0+
DOS 6.0+

DOS 6.2
DOS 6.22
IBM AIX
MACII SYSTEM 7+
MAC (version unknown)
MAC 7
NT (version unknown)
NT 3.51
IRIX/SGI
SGI (version unknown)
SGI INDIGO
SUN (version unknown)
SUN SOLARIS
UNIX (version unknown)
UNIX SHELL
WINDOWS (version unknown)
WINDOWS DDE
WINDOWS 2.0
WINDOWS 2.1
WINDOWS 3.1+
WINDOWS 3.2
WINDOWS 3.11
WINDOWS NT
WINDOWS 95+
WINDOWS 96
WINDOWS 97
WINDOWS FOR PEN COMPUTING
9X/NT

WEB BROWSERS REQUIRED BY OBJECTS IN COLLECTION

Microsoft Internet Explorer (version unknown)
Microsoft Internet Explorer 4.0
Netscape Communicator 4.0
Netscape Communicator 4.5
USMCIB

INVENTORY DATA

Object furnishes enough information to insure full use of the content (i.e.: Readme files?)	Yes: 16 No: 12 Unable to open (lack of required software): 7
Operating system	Windows: 21 DOS: 14

	UNIX: 0 Mac: 0
Object is a document, complex object or other interactive product, or both	Document: 16 Interactive: 14 Both: 5
Metadata adequately describes the object	Yes: 20 No: 16
Data storage requirement per object	Smallest: .02 MB Largest: 10.1 MB Average: 2.21
Duplicated in other media in collection	Yes: 11 No: 24
Average times each has been ordered	34 objects: 3.2 times One object: 358 times
Presence of hyperlinks	Objects in which links were noted: 3 Cases in which these links are broken: 3

B. CD-ROM

SOFTWARE AND FILE TYPES IN COLLECTION

Active Movie 1.0
Active X
Acrobat 2.1
Acrobat 3.0
Acrobat 3.01
Acrobat 4.0
Acrobat Reader
Acrobat Reader 2.1
Acrobat FLE
Acrobat Reader and Search
ANSI-C
Apple Quick Time 2.12
Asymetrix Toolbox II
Authorware 3.01 for Windows
ASCII
AutoCad (DXF)
AVI (MS Audio/Video)
Bit Map
CSS
C++ 4.1+
C++ 7.1+
.DAW
DBase4 ver. 1.5
Fortran Compiler for Program Mugurcs F
FoxPro
GIF
GRF (DPlot)
DPlot 95 (DAT/GRF)
Honeywell Digital Memory Unit (P/N 8509830)
HTML 5
Hypermedia Reader 3.1
IGES
Indeo Audio 2.5
Indeo Video 5.1
Integraph Microstation
JBS file type
JPEG Color
KAware
Lotus Approach
MATLAB
MATLIB (typo: should be MATLAB??)

MPlayer 2
MPG
MS Office
MS Office 97
MS Access (version unknown)
MS Access 2
MS DOS CD-ROM Extensions 2.0+
MS Excel 5.0
MS Excel 7.0
MS Excel 97
MS Excel 97 (SR-1)
MS Excel (version unknown)
MS Front Page
MS Graph
MS Pagemaster
MS Powerpoint
MS Powerpoint v. 8
MS Powerpoint 97
MS Powerpoint
MS Video
MS Word (version unknown)
MS Word 2.0
MS Word 6.0
MS Word 7.0
MS Word 97
MS Word 97 (SR-1)
Movies
NOARL designed s/w created database
NTP file type
Postscript
Postscript.ps
Postscript.eps
Postscript.tex
Quick Time Audio
Quick Time movie
Quick Time 3.0
TAC 3
Textfile
TIFF
Visual Basic 4
Visual C++
WAV
WordPerfect 6
XML

OPERATING SYSTEMS AND VERSIONS IN COLLECTION

IRIX (SGI UNIX) v. 6.2+
IRIX MIRSp50
MS-DOS v. 3.1+
MS-DOS v. 3.3+
MS-DOS v. 5.0
MS-DOS v. 5.0+
MS-DOS v. 5.1
MS-DOS v. 6.0
MS-DOS v. 6.2
MS-DOS (version unknown)
OS2
SGI (IRIX?)
Sun (version unknown)
Sun Solaris 2.5+
Sun Solaris SunPro
Windows 95
Windows 95, v. 4
Windows 98
Windows v. 2
Windows v. 3.1
Windows v. 3.1+
Windows v. 3.11
Windows NT (version unknown)
Windows NT v. 3.5
Windows NT v. 4

DISKETTE INVENTORY DATA

Object furnishes enough information to insure full use of the content (i.e.: Readme files)	Yes: 15 No: 6 Unable to open: 4
Operating system (object may support multiple systems)	Windows: 29 DOS: 1 UNIX: 2 Mac: 8
Object is a document, complex object/interactive product	Document: 13 Complex/interactive: 17*
Metadata adequately describes the object	Yes: 22 No: 23 (Enough metadata is present for customers to determine if the object will work with their systems)
Data storage requirement per object	Smallest: 6 MB Largest: 645 MB Average: 123 MB
Duplicated in other media in collection	Yes: 2 No: 23
Average times each has been ordered	29 objects: 5.6 times
Presence of hyperlinks	Objects in which links were noted: 4 Cases in which these links are broken: 2

*Some of the complex/interactive objects could be considered as documents if internal or web links are ignored.

C. *VIDEOTAPE*

INVENTORY DATA

Number with related items	68 (46%) have related documents or accessories such as 3D glasses
Do related items have same security classification as the tape?	Yes, in all cases
Do all related items have same distribution limits?	Approximately 50%
Format of related items	Furnished as paper documents
All pieces are required to operate the tape	None require all pieces
Tape format	¾ inch: 3 ½ inch VHS: 142
Length of recorded moving image on each tape accession	Total of 143 hours of tape. Range from shortest (3 minutes to longest (12 hours, 40 minutes on 23 tapes) Average length per accession: 37 minutes.
Title slates are included	Both surveyed tapes did
Creation/production credits present	Yes, for both tapes observed
Presence of same content in other formats in collection	Record indicates that a paper equivalent is available: 4 tapes
Image is closed captioned	Approx. 10% of tape collection
Does the content stand alone? Must all pieces be available for one to work?	It is assumed that all parts of the kit are necessary to take full advantage of the content.
Is special equipment required to run or view the tape?	Aside from the one case in which 3D glasses were required, this is not noted in any metadata.